



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

The fragility of opinion formation in a complex world

Medo, Matus ; Mariani, Manuel ; Lü, Linyuan

Abstract: How does the complexity of the world around us affect the reliability of our opinions? Motivated by this question, we quantitatively study an opinion formation mechanism whereby an uninformed observer gradually forms opinions about a world composed of subjects interrelated by a signed network of mutual trust and distrust. We show numerically and analytically that the observer's resulting opinions are highly inconsistent (they tend to be independent of the observer's initial opinions) and unstable (they exhibit wide stochastic variations). Opinion inconsistency and instability increase with the world's complexity, intended as the number of subjects and their interactions. This increase can be prevented by suitably expanding the observer's initial amount of information. Our findings imply that an individual who initially trusts a few credible information sources may end up trusting the deceptive ones even if only a small number of trust relations exist between the credible and deceptive sources.

DOI: <https://doi.org/10.1038/s42005-021-00579-3>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-202728>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Medo, Matus; Mariani, Manuel; Lü, Linyuan (2021). The fragility of opinion formation in a complex world. *Communications Physics*, 4:75.

DOI: <https://doi.org/10.1038/s42005-021-00579-3>

The fragility of opinion formation in a complex world

Matuš Medo^{1,2,3✉}, Manuel S. Mariani^{1,4} & Linyuan Lü^{1,5,6}

How does the complexity of the world around us affect the reliability of our opinions? Motivated by this question, we quantitatively study an opinion formation mechanism whereby an uninformed observer gradually forms opinions about a world composed of subjects interrelated by a signed network of mutual trust and distrust. We show numerically and analytically that the observer's resulting opinions are highly inconsistent (they tend to be independent of the observer's initial opinions) and unstable (they exhibit wide stochastic variations). Opinion inconsistency and instability increase with the world's complexity, intended as the number of subjects and their interactions. This increase can be prevented by suitably expanding the observer's initial amount of information. Our findings imply that an individual who initially trusts a few credible information sources may end up trusting the deceptive ones even if only a small number of trust relations exist between the credible and deceptive sources.

¹Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, People's Republic of China.

²Department of Radiation Oncology, Inselspital, University Hospital of Bern, and University of Bern, Bern, Switzerland. ³Department of Physics, University of Fribourg, Fribourg, Switzerland. ⁴URPP Social Networks, Universität Zürich, Zürich, Switzerland. ⁵Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou, People's Republic of China. ⁶Beijing Computational Science Research Center, Beijing, People's Republic of China. ✉email: matus.medo@unifr.ch

Identifying potential mechanisms behind the formation of opinions in society is vital to understand how polarization emerges in society¹, how misinformation spreads and can be prevented², and how science can be effectively communicated to the public³. Despite recent advances in how opinions propagate in social networks¹, how artificial agents promote low-credibility content in social media^{4,5}, and how rapidly misinformation spreads compared to reliable content^{2,6}, misinformation still thrives in our society. This is well-exemplified by the recent growth of anti-vaccination views and the related existence of anti-vaccination clusters in online social networks⁷. The popularity of unreliable opinions—which is especially dangerous during global emergencies, such as the recent COVID-19 pandemic⁸—calls for a deeper investigation of the possible drivers behind the process whereby individuals form opinions in a society.

Models of opinion formation in social systems have been studied in statistical physics^{9–11}, sociology^{12–16}, and game theory¹⁷. They focus on opinion propagation on a social network of influence and allow us to model phenomena such as the emergence of consensus in the population, for example. Most existing models, however, neglect the potential interconnectedness among topics, which is instead a key aspect of modern social and information environments^{18,19}. These models view opinions of an individual on different topics as results of independent realizations of opinion formation processes, and they neglect direct connections among opinions on distinct topics.

Departing from existing models, we consider a situation where an individual observer forms opinions about a set of subjects interrelated by positive and negative links. The subjects on which the opinions are formed can represent governments, politicians, news media, or other individuals. Connections between subjects can take various forms²⁰. One credible news source can, for example, tend to report in agreement with another credible source (a positive link) whereas a misinformation source can tend to disagree with both credible information sources (negative links). Such systems with signed relations are described by Heider's social balance theory^{21–23} as documented using data on armed conflicts among countries^{24,25} and large-scale social media^{26–28}, for example. Of particular interest is the tendency for the subjects to form two opposing camps^{24,29,30}, although these macroscopic structures are generally imperfect. Two countries, for example, can belong to the same alliance whose members generally have positive relations, yet their mutual relation can be negative due to historic or economic reasons. In science, there are multiple Nobel prize recipients who endorsed unscientific theories³¹.

We study how an observer equipped with prior information about some subjects can use noisy signed relations between subjects divided into two camps to form opinions on the remaining subjects. Forming a reliable opinion in a complex situation

requires effortful reasoning. However, psychological research indicates that humans tend to be rather driven by simple heuristics when forming opinions about complex topics, sometimes reaching opinions that violate basic logic rules^{32,33}. The limitations of our cognition have important consequences. For example, the susceptibility to partisan fake news was recently found to be driven more by “lazy reasoning” than by partisan bias³⁴. For this reason, we focus here on local rules (heuristics) of opinion formation using a signed network of relations.

We find that even a small fraction of misleading links in the relation network among the subjects (e.g., a link of mutual trust between a scientific and low-credibility information source) leads to the resulting opinions that are both inconsistent with the observer's seed opinions and vary significantly between model realizations. We determine analytically the relation between average opinion consistency and the world complexity, represented by the number of subjects, which demonstrates that opinion inconsistency grows as the world complexity increases. This increase can be prevented by suitably increasing the observer's initial number of independent opinions. Although opinion consistency depends on network topology and can be improved by considering a more sophisticated local opinion formation mechanism, our main conclusions are robust to variations of the network topology and the opinion formation mechanism.

Our findings point that even in the absence of social influence, opinion formation in a world composed of many interrelated subjects is inherently fragile. Since subjects may represent co-existing scientific or low-credibility information sources, our model presents a contributing mechanism for how misinformation sources may gain their audience. Our work paves the way to studying strategies to increase the reliability of opinion formation.

Results

Opinion formation model. We consider an individual observer who gradually develops opinions on a world composed of N interrelated subjects (Fig. 1a). The number of subjects represents the complexity of the world. Each opinion is for simplicity assumed to take one of three possible states: no opinion, a positive opinion (trust), or a negative opinion (distrust). The observer's opinions can be formally represented by an N -dimensional opinion vector \mathbf{o} whose element o_i represents the opinion on subject i ; $o_i \in \{-1, 0, 1\}$ corresponds to a negative opinion, no opinion, and a positive opinion, respectively. The subjects form a signed undirected network of relations. These relations are represented by a symmetric $N \times N$ relation matrix whose element R_{ij} represents the trust relation between subjects i and j ; $R_{ij} \in \{-1, 0, 1\}$ corresponds to a negative relation, no relation, and a positive relation, respectively. We emphasize the main difference between this setting and traditional opinion formation models based on propagation on networks of social

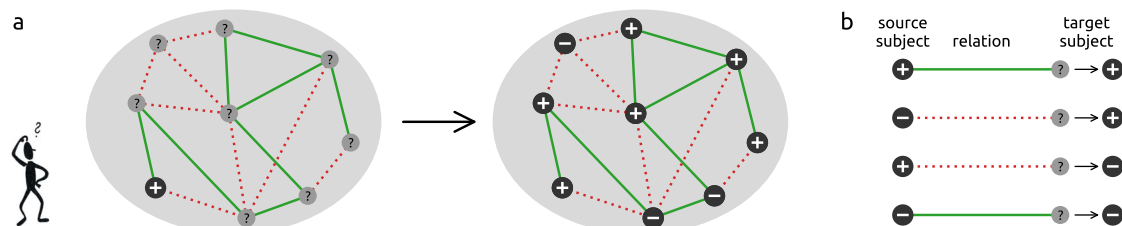


Fig. 1 The opinion formation model. **a** An observer faces a set of subjects that are interconnected by mutual relations of trust (solid green lines) or distrust (dotted red lines). Starting from a small set of seed opinions (here one positive seed opinion marked with + in a black circle) and a world of unknown subjects (gray circles), the observer gradually forms opinions on all subjects. **b** The formed opinion is determined as a product of the opinion on a chosen source subject (one of the seed subjects or any other subject on which an opinion has been already made) and the sign of the relation between the source subject and the target subject. A positive opinion is formed when the source opinion and the relation are both positive or both negative; a negative opinion is formed otherwise.

influence^{9,10,13,14}. In existing models, simulating the opinion formation on N subjects would require running N independent realizations of the opinion formation process, which would miss the interconnectedness among subjects; by contrast, in the proposed approach, the interconnectedness among subjects is naturally encoded in the relation matrix \mathbf{R} .

The observer's opinion formation starts from an initial condition where the observer has an initial opinion on a “seed” subset of subjects, \mathcal{S} (seed opinions). The observer then gradually forms an opinion on each of the remaining subjects via sequential opinion formation, until opinions on all subjects are formed. Once formed, the opinions are not updated (we relax this assumption later). In one step, a target subject i is chosen at random from the pool of subjects with no opinion ($o_i = 0$). The observer then attempts to form an opinion on i . From all subjects j with an opinion ($o_j \neq 0$) that are adjacent to i ($R_{ij} \neq 0$), we choose source subject k at random. The opinion o_i is then set to $o_k R_{ki}$ (see Fig. 1b). As a result, a positive opinion on i is formed if either: (1) the observer has a positive opinion on k and the relation between k and i is positive (“the friends of my friends are also my friends”) or (2) the observer has a negative opinion on k and the relation between k and i is negative (formalizing the ancient proverb “the enemies of my enemies are my friends”). A negative opinion on i is formed otherwise. Note that this mechanism produces a balanced triad consisting of the observer and subjects i and k (in Heider's original sense of heterogeneous triads that can include both individuals as well as objects^{21,35}). The observer then continues with a next subject until opinions on all subjects have been formed. This opinion formation process—which we refer to as the random neighbor rule as it forms opinions using neighboring subjects chosen at random—is purposely simple as it intends to imitate an observer with limited cognitive resources (see ref. ³⁴ for a recent account on susceptibility to fake news driven by “lack of reasoning”). We study a more thorough process (majority rule) below.

The opinion formation outcome is not deterministic (except for the special case when all paths in the subject network are balanced; see Sec. S1 in the Supplementary Information (SI)) as it is influenced by: (1) the order in which subjects are chosen as targets and (2) the choice of the source subject for each target subject. For a given relation network and a set of seed opinions, various resulting opinions are thus possible. Individual realizations of the opinion formation process correspond to a population of independent individuals or, alternatively, various possible “fates” of a single individual. To characterize statistical properties of the resulting opinions, we study outcomes of multiple model realizations. For simulations on synthetic relation networks, we additionally average over various network realizations to remove possible effects of a particular network realization on the results.

Opinion formation simulations on synthetic networks. We now study the opinion formation model on a specific relation network where the subjects form two camps. This scenario is relevant to various real situations^{24,29,30}. The two camps can represent two opposing political parties (such as democrats and republicans), standard news outlets and false news outlets, or scientists and conspiracy theorists, for example. In synthetic networks, each camp consists of $N/2$ subjects. Every subject is connected by signed links with z random subjects, thus creating a regular random network of trust with node degree z . If subjects from the same camp are linked, the sign of their relation is $+1$ with probability $1 - \beta$ and -1 otherwise. Similarly, if subjects from different camps are linked, the sign of their relation is -1 with probability $1 - \beta$ and $+1$ otherwise. Parameter $\beta \in [0, 0.5]$ thus

plays the role of structural noise. As β grows, the negative relations become more common within each camp and positive relations become more common across the camps. When $\beta = 0.5$, the two camps become indistinguishable by definition. The network's level of structural balance^{22,36} is the ratio of the number of balanced triads to all triads in the network. In our case,

$$B = (1 - \beta)^3 + 3(1 - \beta)\beta^2, \quad (1)$$

which corresponds to either all links of a triad (the first term) or one link of a triad (the second term) respecting the two-camp structure, producing a balanced triad as a result. B grows monotonously with β . The equation can be inverted, yielding $\beta = (1 + \sqrt[3]{1 - 2B})/2$, which can be used to write our results in terms of B instead of β .

We assume that the observer has initially a positive opinion on N_S seed subjects from camp 1, and we examine whether the observer ends up with a positive opinion on other subjects from camp 1 and a negative opinion on subjects from camp 2, or not. If the two camps represent scientists and conspiracy theorists, for example, the corresponding practical question is whether an observer who initially trusts a scientist would end up predominantly trusting scientists or conspiracy theorists. Without noise ($\beta = 0$), the opinion formation leads to a definite outcome: A positive opinion on all subjects from camp 1 and a negative opinion on all subjects from camp 2. In such a case, we say that the opinions are perfectly consistent with the underlying two-camp structure of the relation network among the subjects. Opinion consistency of a resulting opinion vector, \mathbf{o} , can be measured as

$$C(\mathbf{o}, \mathbf{T}) = \frac{1}{N - N_S} \sum_{j \notin \mathcal{S}} o_j T_j \quad (2)$$

where \mathcal{S} is the set of seed subjects and \mathbf{T} is the vector representing the ground-truth structure of the relation network (in our case, $T_j = 1$ for j from camp 1 and $T_j = -1$ for j from camp 2). If the observer's opinions are chosen at random, the resulting consistency is zero on average. A zero or small consistency value thus indicates that the observer's opinions are independent of the seed opinion and thus inconsistent with the two-camp structure of the relation network. Negative consistency is also possible: The observer starts with a positive opinion on subjects from camp 1 but ends with more positive opinions in camp 2 than in camp 1.

Knowing that opinion consistency is one in the absence of noise, how does it change as the noise parameter β grows? Numerical simulations for a set of 100 subjects and one seed opinion show (Fig. 2a) that opinion consistency decreases rapidly with β . Indeed, if the relationship between consistency and noise was linear, we would have expected $C_0(\beta) := 1 - 2\beta$ which starts at one when $\beta = 0$ and reaches zero when $\beta = 0.5$ as the two camps then cannot be distinguished by definition. By contrast, we observe a substantially faster decay of the mean consistency $\mu_C(\beta)$. In addition, the consistency values vary strongly between model realizations. For $\beta = 0.02$, for example, mean consistency is only 0.80 and there are model realizations with consistency below 0.54 and above 0.97 (the 10th and 90th percentile, respectively, of the obtained consistency values for $z = 4$). This means that even when the noise is small, some sets of formed opinions are in a dramatic disagreement with the observer's seed opinion. To appreciate the level of noise in real data, Moore²⁴ reported that 80% of triads among middle East countries are balanced. Equation (1) shows that such a level of structural balance is achieved at $\beta \approx 0.08$ in our two-camp networks. In Fig. 2b, mean opinion consistency at $\beta = 0.08$ is as low as 0.42 (for $z = 10$). These results confirm our initial hypothesis that a realistic level of

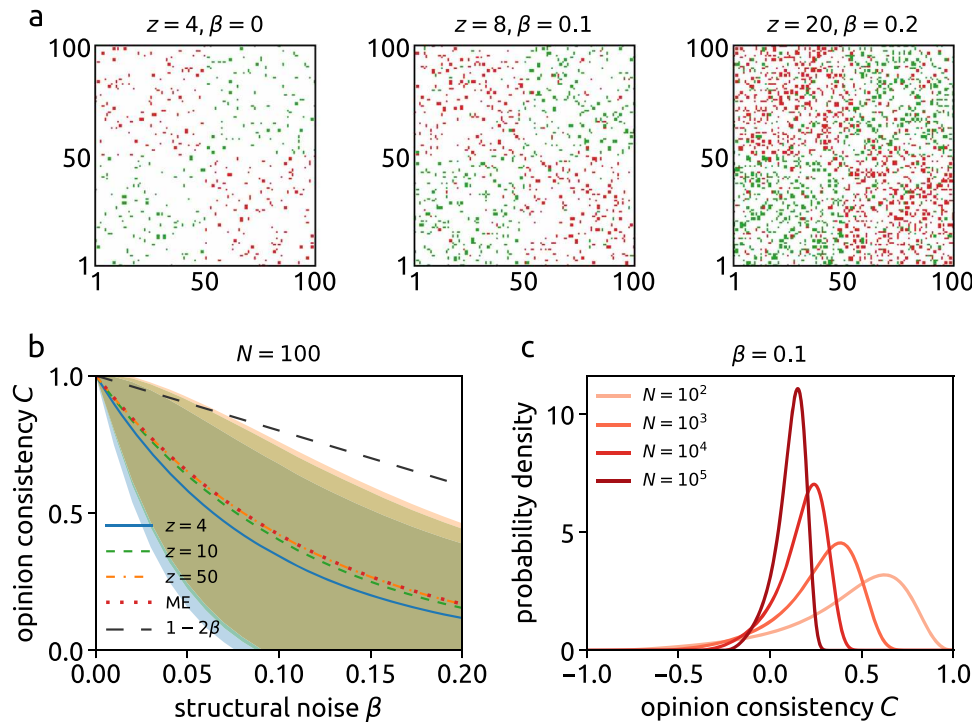


Fig. 2 Opinion formation on random relation networks with a two-camp structure. **a** Examples of adjacency matrices of regular random networks with a two-camp structure for 100 subjects and various values of node degree, z , and structural noise, β . Subjects 1–50 are from camp 1 and subjects 51–100 are from camp 2. The green, red, and white points represent positive, negative, and absent relations, respectively. **b** Resulting opinion consistency for $N = 100$ subjects, one seed opinion, and various values of z . The lines show mean values and the shaded regions show the 10th–90th percentile ranges (both computed from 1000 model realizations for each of 1000 realizations of the relation network). The dotted line labeled ME shows the mean consistency computed by numerically solving the master equation given by Eq. (3). The dark dashed line shows the expected consistency if its relationship with the level of noise was linear [$C_0(\beta) = 1 - 2\beta$]. These results demonstrate that consistency decreases quickly with structural noise. **c** The consistency distributions obtained using Eq. (3) for $\beta = 0.1$, one seed subject, and a growing number of subjects, N . As N increases, the distributions shrink and their peaks shift toward zero.

noise leads to the adoption of a large fraction of opinions that do not align with the observer’s initial opinion.

Master equation for opinion consistency and its solution. The opinion formation with the two-camp structure can be studied analytically under the assumption of homogeneous mixing³⁷. It is advantageous to study the problem in terms of the number of formed opinions, n , and the number of consistent opinions, c (that is, the opinions that are consistent with the seed opinions and the two-camp structure). By rewriting the sum $\sum_{j \neq i} o_j T_j$ in Eq. (2) as $2c - N + N_S$, we obtain opinion consistency as $C = (2c - N + N_S)/(N - N_S)$.

When the observer forms a new opinion, n increases by one and c either increases by one (if the new opinion is consistent) or remains constant. We introduce the probability distribution of c when n opinions have been formed, $P(c; n)$, for which the master equation (see “Methods” for the derivation) has the form

$$P(c; n) = P(c-1; n-1) \frac{c(1-2\beta) + \beta(n+1) - 1}{n-1} + P(c; n-1) \left[1 - \beta - \frac{c(1-2\beta)}{n-1} \right] \quad (3)$$

The initial condition $P(N_S; N_S) = 1$ represents that all N_S seed opinions are consistent. Equation (3) can be solved numerically and the obtained solution $P(c; n)$ can be used to compute the corresponding mean opinion consistency. The numerical solution agrees well with the model simulations (Fig. 2b), in particular when the relation network is not sparse ($z \gtrsim 10$).

Equation (3) allows us to analytically study the dependence between opinion consistency and the world complexity, represented

by the number of subjects, N . A surprising finding is that as the number of subjects increases, the distribution of C obtained by solving Eq. (3), $P(C)$, does not approach a well-defined limit distribution, but instead steadily shifts toward $C = 0$ and becomes narrower in the process (Fig. 2c). We study this behavior by computing the mean opinion consistency, $\mu_C(N)$, and the standard deviation of consistency, $\sigma_C(N)$ (see Sec. S2 in the SI for details and additional analytical results).

Multiplying Eq. (3) with c and summing it over $c = N_S, \dots, N$ yields the recurrence equation

$$\langle c(n) \rangle = \frac{n-2\beta}{n-1} \langle c(n-1) \rangle + \beta \quad (4)$$

with the initial condition $\langle c(N_S) \rangle = N_S$ (the seed opinions are assumed to be correct). This recurrence equation can be solved in general, yielding

$$\langle c(N) \rangle = \frac{1}{2} \left[N + \frac{\Gamma(N+1-2\beta)\Gamma(N_S+1)}{\Gamma(N_S+1-2\beta)\Gamma(N)} \right]. \quad (5)$$

For $N_S = 1$, the corresponding mean consistency is

$$\mu_C(N) = \left[\frac{\Gamma(N+1-2\beta)}{\Gamma(2-2\beta)\Gamma(N)} - 1 \right] / (N-1) \quad (6)$$

which simplifies to $\mu_C(N) = 1$ when $\beta = 0$, as expected. The leading contribution of Eq. (6) is

$$\mu_C(N) = N^{-2\beta} / \Gamma(2-2\beta) + O(1/N). \quad (7)$$

This shows that when $\beta > 0$, the mean opinion consistency vanishes in the limit $N \rightarrow \infty$. The leading-term contribution to

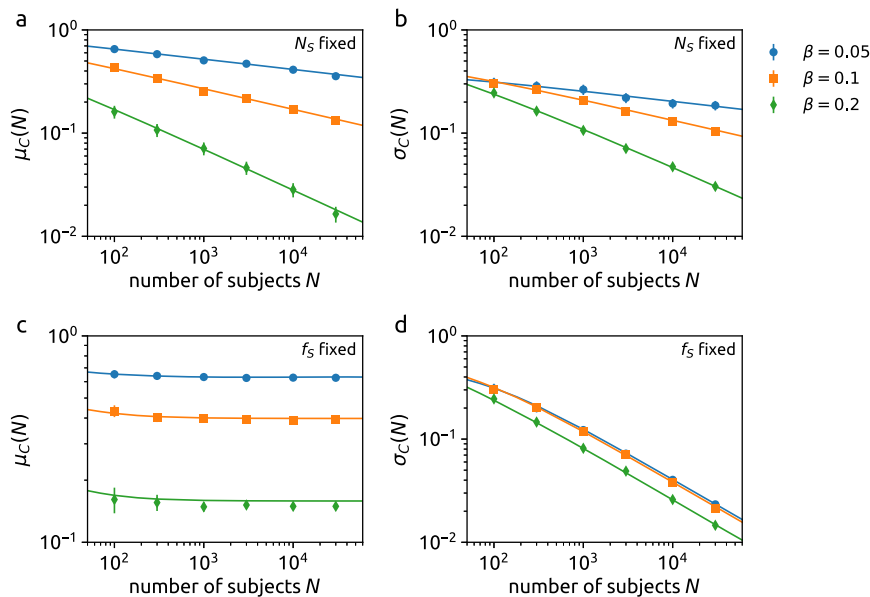


Fig. 3 Scaling of opinion consistency in synthetic worlds. Mean opinion consistency, μ_C , and the standard deviation of consistency, σ_C , as functions of the number of subjects, N , for regular random networks with node degree $z = 50$ and various levels of structural noise, β . The symbols and error bars show the mean and three times the standard error of the mean, respectively, obtained by running the model on 1000 network realizations. The solid lines show $\mu_C(N)$ and $\sigma_C(N)$ obtained by solving the master equation (Eq. (6) and Eq. (S6) in the SI). **a, b** Results for a single seed opinion. In this scenario, both μ_C and σ_C converge to zero as N grows: There is a tension between opinion consistency and the world's complexity. **c, d** Results for a fixed fraction of seed opinions, f_S , when the number of seed opinions is $N_S = f_S N$; here $f_S = 0.01$. In this scenario, the tension between consistency and complexity is prevented: μ_C converges to $f_S^{2\beta}$ (**c**) and σ_C converges to zero (**d**).

$\sigma_C(N)$ is also proportional to $N^{-2\beta}$ when $\beta \leq 1/4$. When $\beta > 1/4$, the leading term becomes proportional to $N^{-1/2}$. These analytic results agree with numerical simulations of the model (Fig. 3a, b and Fig. S1 in the SI).

The behavior demonstrated by Figs. 2b and 3, and supported by the analytic solution above, has important consequences. It shows that as the world complexity increases, the formed opinions become on average less consistent with the seed opinions and the two-camp structure of the subject network. Crucially, the opinion consistency is zero in the limit of an infinite number of subjects for any positive level of noise, β , in the subject relation network: In the limit of an infinite-complexity world, even a tiny amount of noise is enough to nullify opinion consistency.

The convergence of opinion consistency to zero as $N \rightarrow \infty$ can be avoided if the number of seed opinions grows linearly with N so that the fraction of seed opinions remains constant. Assuming that $N_S = f_S N$, Eq. (5) can be used to show that the mean consistency approaches to

$$\mu_C = f_S^{2\beta} \quad (8)$$

in the limit $N \rightarrow \infty$ and the standard deviation of consistency vanishes as $1/\sqrt{N}$ (see Sec. S2.2 in the SI). This scaling relation determines the necessary proportion of seed opinions, f_S , needed to achieve a desired opinion consistency, μ_C , for given β . These results are confirmed by numerical simulations shown in Fig. 3c, d and Fig. S2 in the SI. Despite having a positive limit value, opinion consistency still decreases quickly with noise in the relation network when f_S is small.

While our findings hold qualitatively when a different topology of the relation network is used, the mean opinion consistency values are heavily affected by the network topology (see Fig. 4). We run the opinion formation model on a growing preferential attachment network, a configuration model (CN) network with a power-law degree distribution, and Watts–Strogatz networks with

various values of the rewiring probability, p_r (see Sec. S3 in the SI for details on the network construction and additional simulation results). We find that networks with broad degree distributions lead to higher opinion consistency which decays with N slower (see Figs. S3 and S4 in the SI) than in the previously studied random networks. By contrast, Watts–Strogatz networks yield lower opinion consistency which further decreases as the networks become more regular through lowering the rewiring probability, p_r .

Opinion formation using the majority rule. The results described above hold for the opinion formation model where a random neighbor of a target subject is chosen as the reference. We chose this model to study the consequences of a cognitively easy opinion formation model. At this stage, one might object that the observed sensitivity of opinion consistency to noise might be because each formed opinion directly relies on only one previously formed opinion, and it might disappear if the observer incorporates the information from more neighbors before forming an opinion. To rule out this potential argument, we investigate a model where all neighbors of a target subject are considered before forming the opinion. Denote the numbers of neighbors leading to the adoption of a positive and a negative opinion (determined as in Fig. 1b) as n_P and n_N , respectively. If $n_P > n_N$, the observer forms a positive opinion. If $n_N > n_P$, the observer forms a negative opinion. If $n_P = n_N$, a random opinion is formed. We refer to this as the majority opinion formation rule. It is more demanding than the original random neighbor rule based on choosing a random neighbor as it assumes that the observer carefully collects all evidence for forming an opinion on a target subject. The majority rule is nevertheless still a local rule as it only considers direct neighbors of a target subject.

Using the majority rule, a scaling analogous to Fig. 3 can be observed (Fig. 5a, b) when the number of seed opinions is fixed. The important difference is that the scaling exponent now depends on both β and z whereas a higher mean degree, z ,

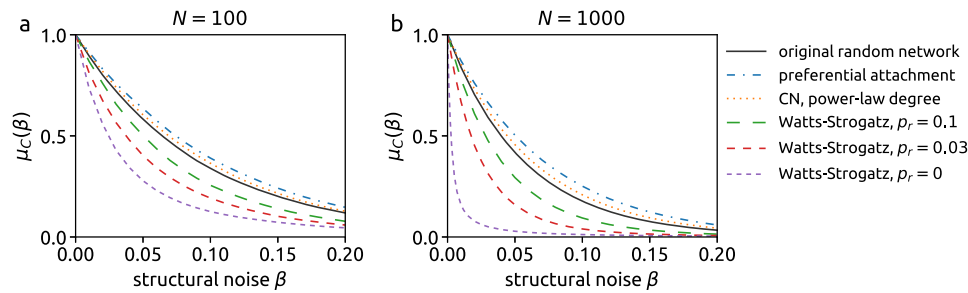


Fig. 4 Opinion consistency for various topologies of synthetic relation networks. The dependence of mean opinion consistency, μ_C , on structural noise, β , for one seed opinion and different numbers of subjects: **a** $N = 100$ and **b** $N = 1000$. The results are averaged over 1000 model realizations on each of 1000 network realizations, mean degree $z \approx 4$ for each network topology. The error bars are too small to be shown. CN and p_r stand for the configuration model and the rewiring probability in the Watts-Strogatz model, respectively.

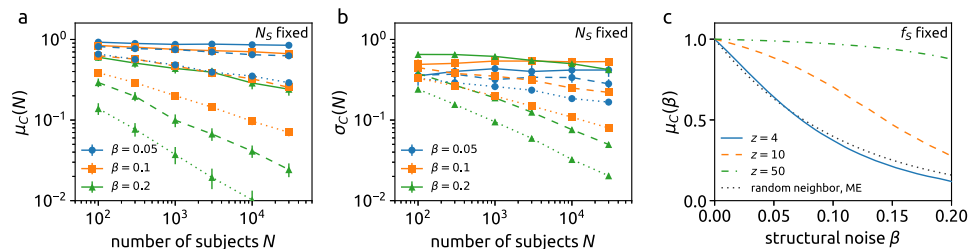


Fig. 5 Scaling of opinion consistency achieved by the majority rule in synthetic worlds. **a, b** The dependencies of μ_C and σ_C on the number of subjects, N , are affected by both structural noise, β , and mean degree, z (the solid, dashed, and dotted lines are used for z equal 50, 10, and 4, respectively). One seed opinion is used ($N_s = 1$). The symbols and error bars show the mean and three times the standard error of the mean, respectively, obtained by running the model on 1000 realizations of regular random networks. **c** The dependence of mean opinion consistency, μ_C , on structural noise, β , for a fixed fraction of seed opinions, $f_s = 0.01$, $N = 10,000$, and $N_s = f_s N$. The results are averaged over model realizations on 1000 regular random networks, the error bars are too small to be shown. The dotted line shows the analytical result $\mu_C = f_s^2\beta$ for the random neighbor rule when the same fixed fraction of seed opinions, $f_s = 0.01$, is used.

generally leads to $\mu_C(N)$ and $\sigma_C(N)$ decaying slower with N . Except for the smallest used noise and the highest used degree ($\beta = 0.05$ and $z = 50$), all slopes of linear fits between $\ln \mu_C(N)$ and $\ln N$ are significantly positive. Since the majority rule does not lend itself to analytical computation, whether the limit of $\mu_C(N)$ is indeed positive when β is sufficiently small and z is sufficiently high remains an open question. When the fraction of seed opinions, f_s , is fixed, $\mu_C(N)$ converges to a positive value and $\sigma_C(N)$ vanishes as N grows (Fig. S5 in the SI). The scaling of the majority rule and the random neighbor rule is thus qualitatively the same when f_s is fixed. Figure 5c shows mean opinion consistency as a function of structural noise for fixed f_s when the number of subjects is large. We see that when the network density is low ($z = 4$), the majority rule achieves results that are comparable with those of the random neighbor rule. When z increases, the majority rule leads to significantly more consistent opinions than the random neighbor rule. It has to be noted, though, that when z is large, the cost for the observer to collect and analyze all information for opinion formation is large too.

The effect of opinion updating on opinion consistency. To focus on the opinion formation mechanism alone, the original model assumes that the formed opinions cannot be changed anymore. In real life, however, we occasionally re-evaluate our positions. This re-evaluation can be included in the model by allowing the opinions to be updated as it is, in fact, common in the opinion formation literature^{17,38,39}.

To incorporate an updating mechanism while preserving the spirit of the model, we assume that in the first phase, all opinions are formed using the original model. In the second phase, target subjects for opinion updating are repeatedly chosen from all subjects at random. Opinions are updated in the same way as they

have been formed in the first phase, i.e., using the random neighbor rule or the majority rule. Simulation results (see Fig. S6 in the SI) show that the effect of opinion updating markedly differs between the two opinion formation rules. When the random neighbor rule is used, opinion consistency substantially decreases with the number of updated opinions. The reason for this decrease lies in the opinion formation rule which relies on one randomly chosen neighbor of the target subject. Noise in the relation network thus continues to accumulate through updating which thus lowers opinion consistency. By contrast, opinion consistency substantially increases with the number of updated opinions when the majority rule is used. Thanks to its thorough (albeit local) consideration of opinions on neighbors of the target subject, the majority rule is capable of identifying and correcting imbalanced triads in the relation network, and thus increase opinion consistency.

Opinion formation simulations on real networks. The trust consistency metric requires information on the ground truth structure of the relation network (such as the assignment of subjects to one of the two camps in the case of a two-camp structure). Before analyzing empirical data, we aim to introduce a proxy for opinion consistency that does not require such information which is typically not available for real data. To this end, we introduce opinion stability, S , which measures the extent to which elements of the opinion vector are the same in independent realizations of the opinion formation model (see “Methods” for the definition). If an opinion on a given subject always ends up positive (or always negative), it is a sign of a robust opinion and it contributes positively to opinion stability. Small opinion stability indicates that the opinion formation outcomes are highly volatile and, in turn, they do not comply with the division of subjects in

camps in synthetic networks. It can be shown that when the relation network's level of structural balance is one, opinion stability is one as well.

In synthetic worlds, the opinion stability metric behaves as required when the relation network is sufficiently dense ($z \gtrsim 10$): $S = 1$ in synthetic networks when $\beta = 0$ and S is close to zero when $\beta = 0.5$ (see Sec. S4 in the SI). In fact, the values of opinion consistency and opinion stability are nearly the same for all β values. The main reason for this agreement between stability, S , and consistency, C , is that high opinion consistency can be only achieved when the opinions in question are the same in all model realizations which in turn leads to high opinion stability. Crucially, opinion stability vanishes as the number of subjects grows to infinity similarly as we have seen it for opinion consistency (see Fig. S9 in the SI).

Equipped with the opinion stability metric, we can assess opinion formation in empirical worlds represented by empirical signed networks. We first use signed networks derived from United Nations General Assembly (UNGA) votes in individual sessions, where countries that vote similarly are connected with positive links and countries that vote differently are connected with negative links (see “Methods” for the data description). Figure 6a shows part of the network corresponding to the latest completed UNGA session 74 (2019–2020). The loop is unbalanced as the product of its link weights is -1 . As a result, the outcome of opinion formation using the random neighbor rule is not deterministic: Assuming a positive seed opinion on Italy, the formed opinion on Russia is negative if it is done using the path ITA-FRA-RUS or positive if it is done using the path ITA-USA-RUS. This outcome variability manifests itself in Fig. 6b, c where two realizations of the random neighbor rule are shown to considerably differ despite a high level of structural balance of the respective UNGA network (in this case, $B = 0.86$). This agrees with our results in Fig. 2b where opinion consistency decreases quickly with β and displays large fluctuations. Finally, Fig. 6d–f shows that the majority rule yields substantially more stable opinions and that the stability difference between the random neighbor rule and

the majority rule tends to grow as the level of structural balance decreases.

The number of nodes in the UNGA datasets is limited by the number of countries participating in the assembly's voting (the number of nodes grows from 53 in the first assembly to 191 in the 74th). To be able to observe the scaling of opinion stability similar to the scaling of opinion consistency in synthetic data (Fig. 3), we thus use signed trust networks from two popular online services: Slashdot^{40,41} and Epinions⁴¹ (see “Methods” for the data description). Note that while Slashdot and Epinions are social networks, our model still differs from classical models of opinion formation on social networks as it concerns opinion formation of an observer, not opinion formation of each individual member of the social network. Nodes in the given social networks represent interconnected subjects on which opinions are formed.

We create multiple subsets of each network with progressively increasing numbers of nodes (see “Methods” for details). We find that a stability–complexity tension is present also in the real worlds (Fig. 7a, b): Opinion stability consistently decreases with the number of subjects (see Fig. S12 in the SI for detailed results showing the mutual effect of the network size and the level of structural balance). The fitted scaling exponents are 0.40 and 0.19 for Slashdot and Epinions, respectively (see Fig. S13 in the SI). Note that these values cannot be directly compared with the scaling exponent 2β that we derived for opinion consistency in random networks as Slashdot and Epinions networks are manifestly non-random. Building on the understanding that we gained by analyzing simulations on synthetic worlds, we can conclude that the levels of noise in the two real relation networks are high, which makes opinion formation using the random neighbor rule unreliable.

We finally study the scaling of the majority rule on real networks (Fig. 7c, d). In agreement with previous results, we observe that the resulting opinion stability is almost always higher than achieved by the random neighbor rule and the difference generally grows as the network's level of structural balance decreases (Fig. S14 in the SI). As N increases, the average opinion stability still vanishes with N , albeit slower than is the case for the

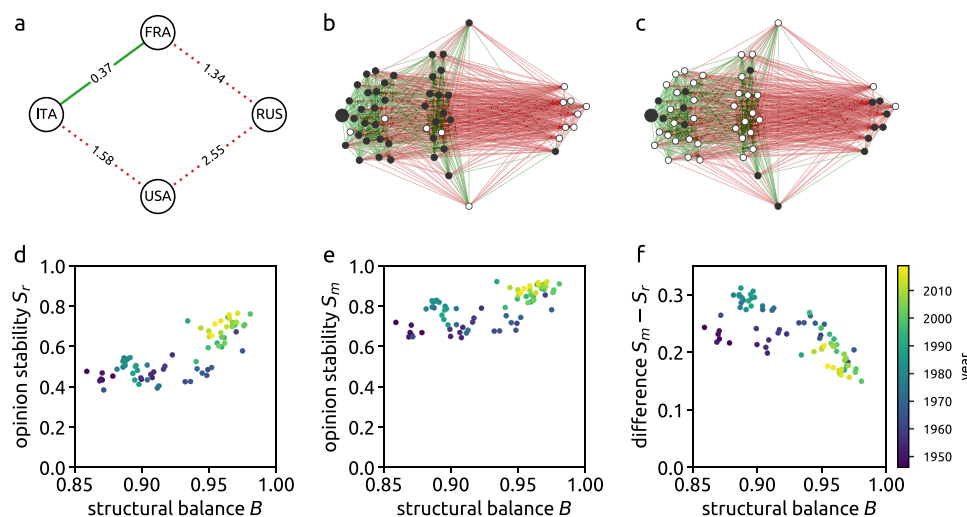


Fig. 6 Opinion stability for the United Nations General Assembly voting data. **a** A sample unbalanced subgraph of the last session's dataset. Links are labeled with absolute differences between the states' ideal point positions estimated in ref. ⁴⁴. Absolute differences below the 33th percentile and above the 67th percentile are represented as positive (solid green) and negative (dotted red) links, respectively. **b, c** Resulting opinions in two different realizations of the random neighbor rule on the first session's dataset. Black and white symbols indicate positive and negative opinions, respectively. The seed opinion (indicated with a larger marker) is the same in both realizations. **d–f** Opinion stability achieved by the random neighbor rule (**d**) and the majority rule (**e**) plotted against the level of structural balance in input relation networks corresponding to 73 individual UN sessions (symbol color indicates the session year). Panel **f** shows the difference between the two rules.

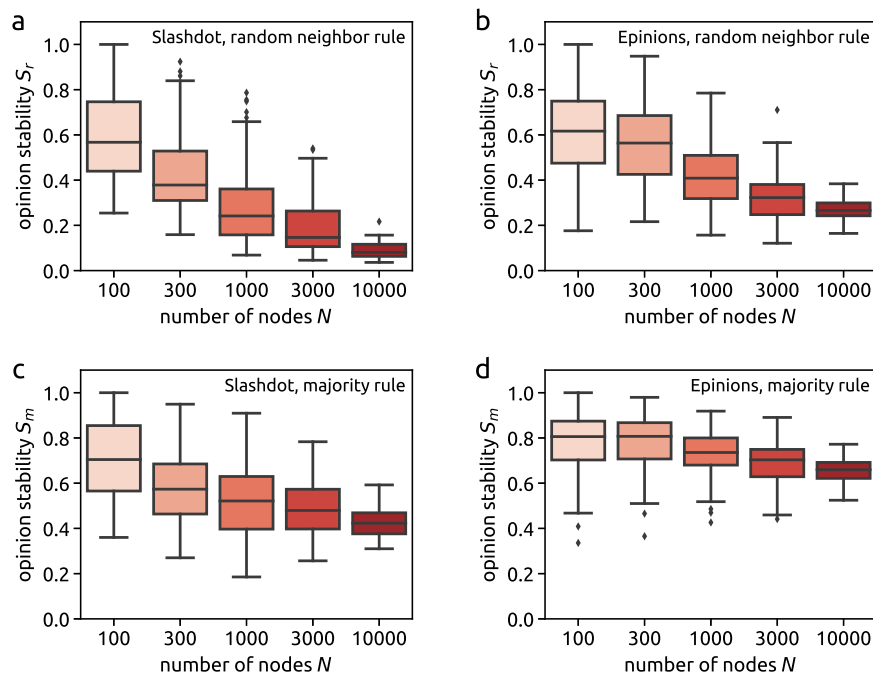


Fig. 7 Scaling of opinion stability in empirical worlds. Opinion stability achieved in subsets of progressively increasing size by the random neighbor rule (a, b) and the majority rule (c, d). Subsets have been obtained by sampling from corresponding empirical signed networks. The boxplots summarize opinion stability values obtained for 100 independent seed opinions (boxes show quartiles of the data and whiskers extend to 1.5 of the inter-quartile range; outlier data points beyond the whisker range are shown individually). In parallel with the results on synthetic relation networks, opinion stability decreases with N (opinion stability–world complexity tension). The decrease is slower for the majority rule than for the random neighbor rule.

random neighbor rule. The fitted scaling exponents are 0.09 and 0.05 for Slashdot and Epinions, respectively (Fig. S15 in the SI). Despite the analyzed real network subsets being rather dense (Fig. S10 in the SI), the opinion stability–world complexity tension is not averted by the majority rule. We can thus again conclude that the majority rule alone does not solve the fundamental problem identified by our work: The formed opinions become progressively less reliable as the system size grows.

Conclusion

We demonstrated that the process whereby an individual sequentially forms opinions on a network of interrelated subjects is inherently fragile: Even a small amount of noise leads to inherently fragile (i.e., inconsistent and unstable) outcomes. This suggests that to prevent the spreading of misinformation in large-scale online systems, it is paramount that there exist no trust links from credible to low-credibility sources of information. If a tiny fraction of such misleading links exists, an observer who starts trusting credible information sources may end up trusting a substantial number of low-credibility sources. If the same happens for a large number of observers, a cluster of misinformed individuals (such as anti-vaccination clusters⁷) can thrive.

The more complex our world, the more fragile the process—there is a tension between opinion consistency/stability and the world’s complexity. An individual observer can compensate for this increasing fragility by forming an independent (i.e., not derived from the trust network) initial opinion on a larger number of subjects, before resorting to the trust network to form an opinion on the remaining subjects. For example, a person forming opinions about a set of interconnected websites—some of them based on scientific content, some of them promoting conspiracy material—can increase the opinion consistency/stability by first carefully evaluating the trustworthiness of a large number of websites, and only then relying on relations between already

trusted/distrusted websites to form opinions on the remaining ones. This suggests that policies to increase the consistency/stability of collective opinions may aim to promote the formation of individuals’ independent opinion about a substantial number of subjects. Within the studied framework, the majority rule yields better results, yet: (1) the majority rule is more laborious than the random neighbor rule as it assumes collecting all direct evidence on each target node, (2) opinion consistency and stability under the majority rule still vanish when the network is not sufficiently dense and the noise is not sufficiently low.

From the observer’s standpoint, our work focuses on little cognitively demanding opinion formation mechanisms, which opens the way to studying more sophisticated mechanisms, and understanding the trade-off between the robustness of the resulting opinions and the observer’s cognitive costs required to form the opinions. In the real world, additional influences—social influence and mass media, in particular—and heuristics are likely to be at work, and a high level of heterogeneity across observers is expected. Our framework could be integrated with that introduced in ref. ⁴², for example, to study whether these additional influences further increase or mitigate the fragility of the individuals’ opinion formation process. More sophisticated opinion formation models and their calibration on empirical data hold promise to shed light on this process which is fundamental for our interconnected and information-driven societies.

Methods

Opinion consistency. When the ground truth division of subjects in camps is known, we compute the consistency of a given opinion vector \mathbf{o} with the ground truth assignment \mathbf{T} using Eq. (2). The seed opinions are excluded from the computation of consistency as these opinions are consistent by construction. The consistency values range from -1 for the opinion vector that totally disagrees with the ground truth (for positive seed opinions on subjects from camp 1, for example, the observer has negative opinions on all remaining subjects from camp 1 and positive opinions on all subjects from camp 2) to $+1$ for the opinion vector that perfectly matches the ground truth camps and the seed opinions. The consistency of a random opinion vector is zero on average. Note that we assume here that

opinions have been eventually formed on all subjects, which is the case for all simulations presented here.

In numerical simulations, we average over independent model realizations on multiple realizations of the synthetic two-camp networks to estimate the mean opinion consistency. In Fig. 2, we complement the mean with the 10th–90th percentile range of the consistency values. In Fig. 3, we assess the uncertainty of mean consistency using the standard error of the mean; the displayed error bars are three times of that.

Master equation solution. The proposed opinion formation model using the random neighbor rule can be studied analytically for the two-camp relation network. We use the number of formed opinions, n , and the number of “consistent” opinions, c , as the variables to describe the process. Initially, $n = N_S$ and $c = N_S$, because all seed opinions are assumed to be consistent (they are all positive opinions on subjects from camp 1 or, more generally, positive opinions on subjects from camp 1 and negative opinions on subjects from camp 2). We introduce the probability distribution of c consistent opinions after n opinions are formed, $P(c; n)$, with normalization $\sum_c P(c; n) = 1$. The initial condition is $P(N_S; N_S) = 1$ in line with the description above. To find $P(c; n)$ for $n > 1$, we write the general master equation

$$P(c; n) = P(c-1; n-1)W(c-1 \rightarrow c; n-1) + P(c; n-1)[1 - W(c \rightarrow c+1; n-1)]. \quad (9)$$

The transition probability $W(c-1 \rightarrow c; n-1)$ corresponds to forming a consistent opinion in a situation when $n-1$ opinions have been formed, of which $c-1$ are correct.

If the target subject (on which the opinion is to be formed) is from camp 1, $W(c-1 \rightarrow c; n-1)$ is the probability that the observer decides to form a positive opinion on the target subject. Assume now that there are t_1 trusted (i.e., with a positive opinion) subjects from camp 1, d_1 distrusted (i.e., with a negative opinion) subjects from camp 1, t_2 trusted subjects from camp 2, and d_2 distrusted subjects from camp 2. The probability of forming a positive opinion on the target node is $n_P/(n_P + n_N)$ where n_P and n_N are the numbers of neighbors of the target subject that—when chosen—would result in forming positive and negative opinions, respectively. The expected value of n_P is proportional to

$$t_1(1-\beta) + d_1\beta + t_2\beta + d_2(1-\beta) = (c-1)(1-\beta) + (n-c)\beta$$

where we used that $t_1 + d_2 = c-1$ and $t_1 + d_1 + t_2 + d_2 = n-1$. Note that it is the random structure of the relation network that allowed us to write a simple expression. Similarly, n_N is proportional to

$$t_1\beta + d_1(1-\beta) + t_2(1-\beta) + d_2\beta = (c-1)\beta + (n-c)(1-\beta).$$

Taken together, these formulas give us the transition probability $W(c-1 \rightarrow c; n-1) = n_P/(n_P + n_N) = [(c-1)\beta + \beta(n-1)]/[(c-1)\beta + \beta(n-1)]$. The form of $W(c-1 \rightarrow c; n-1)$ is the same when the target subject is from camp 2. By plugging this $W(c-1 \rightarrow c; n-1)$ in Eq. (9), we obtain the master equation Eq. (3) which describes how $P(c; n)$ changes as n grows.

Note that the fact that $W(c-1 \rightarrow c; n-1) = n_P/(n_P + n_N)$ implies that the same master equation—and thus the same opinion formation model—is obtained by a seemingly more thorough observer who first evaluates all neighbors of the target subject and counts the number of subjects whose choice would result in forming a positive and a negative opinion, n_P and n_N , respectively. Based on n_P and n_N , the opinion on the target subject can be formed in a probabilistic manner: Positive with probability $n_P/(n_P + n_N)$ and negative otherwise. The outcome is thus the same as choosing one neighbor of the target opinion at random and forming the opinion accordingly.

Opinion stability. Opinion consistency assumes that the ground truth division of subjects in camps is known but that is not the case for most real datasets. To overcome this difficulty, we introduce another metric to assess the formed opinions, opinion stability. For a given relation network, we fix the opinions on subject i and use R independent model realizations to compute the average opinion \bar{o}_j for all other subjects. If the formed opinions on subject j are stable, they are the same in all or most realizations and the value \bar{o}_j is thus close to $+1$ or -1 . By contrast, volatile formed opinions result in \bar{o}_j close to zero. We then compute the average opinion stability with respect to the seed subject i as

$$S' = \frac{1}{N-1} \sum_{j \neq i} |\bar{o}_j| \quad (10)$$

where the absolute value reflects the fact that both $\bar{o}_j = 1$ and $\bar{o}_j = -1$ are signs of stable opinions on subject j . Note that the seed opinion is again excluded from the summation. Opinion stability is $S'_i = 1$ when all realizations yield the same opinion on i . For random opinions, however, the stability is not zero due to the absolute value in Eq. (10) which is never negative. In that case, \bar{o}_j follows the normal distribution with zero mean and standard deviation $1/\sqrt{R}$. It can be shown that the mean of $|\bar{o}_j|$ is $\sqrt{2/(R\pi)}$ which represents the expected opinion stability of a

random trust vector. We thus transform Eq. (10) as

$$S_i = \frac{S' - \sqrt{2/(R\pi)}}{1 - \sqrt{2/(R\pi)}} \quad (11)$$

to obtain the final formula for opinion stability. Its values range from zero, on average, when opinions on all subjects are random to one when opinions on all subjects are the same in all model realizations. Individual S_i values can be used to characterize the stability of opinions based on a seed opinion on subject i or aggregated to represent the overall opinion stability. In simulations on synthetic relation networks, we use 1000 independent network realizations and compute opinion stability for a randomly chosen node. In simulations on real relation networks, we present opinion stability results for 100 nodes chosen at random. See Fig. S7 in the SI for a direct comparison between opinion consistency and opinion stability.

Real datasets. We test the opinion formation model on three distinct real datasets.

The UNGA dataset is derived from votes by countries at United Nations General Assemblies⁴³. We use the state ideal point positions in one dimension estimated in ref. 44 from the voting data to quantify “state positions toward the US-led liberal order”. The dataset contains all 74 general assemblies held in the years 1946–2020; assembly 19 is ignored because of faulty data. For each assembly, estimated state positions x_i can be directly translated in distances $|x_i - x_j|$ between the states. We generate one signed network for each general assembly by first removing all countries with less than 20 votes (up to seven countries have been removed in one session) and then representing state distances below the 33th percentile (for the given general assembly) as positive links and state distances above the 67th percentile as negative links, saving the network’s giant component. The numbers of nodes and links in the network increase progressively from 53 and 919, respectively, in the first general assembly to 191 and 12,096, respectively, in the 74th. The numbers of positive and negative links are identical by construction in each network; the level of structural balance ranges from 0.86 to 0.98.

The Slashdot dataset⁴⁵ represents the social network website social network where the users can tag each other as friends or foes^{40,41}. While the original network is not symmetric, we represent it as symmetric, neglecting the mutual links whose signs do not agree (less than 1% of all links) and finally keeping only the giant component. The resulting Slashdot network comprises 82,052 nodes and 498,527 signed links. The fraction of negative links is 0.236 and the network’s level of structural balance is $B = 0.867$.

The Epinions dataset⁴⁵ represents the social trust network of the website’s users⁴¹. After the same processing as we apply to the Slashdot data, the resulting Epinions network comprises 119,070 nodes and 701,569 signed links. The fraction of negative links is 0.168 and the network’s level of structural balance is $B = 0.905$.

To study the dependence of results on the network size, we created small subsets of the large Slashdot and Epinions networks by choosing a random node and gradually including its nearest neighbors, second-nearest neighbors, and so on, until a target number of nodes is reached. We created 100 independent networks for each network size, each of them starting from a node chosen at random. An alternative construction by choosing a given number of nodes or links at random would produce very sparse networks whose sparsity would directly impact the opinion formation process (see Fig. S8 in the SI). The created subsets, as well as individual UNGA datasets, show levels of structural balance that are considerably higher than levels of structural balance in their randomized counterparts (Fig. S11 in the SI). These empirical datasets are thus suitable testbeds for our opinion formation model. Results obtained on empirical datasets are shown in Figs. 6, 7, and in Sec. S5 in the SI.

Data availability

The UNGA dataset analyzed during the current study is available in the Harvard Dataverse repository, <https://doi.org/10.7910/DVN/LEJUQZ>. The Slashdot dataset analyzed during the current study is available in the Stanford Network Analysis Project repository, <http://snap.stanford.edu/data/soc-sign-Slashdot090221.html>. The Epinions dataset analyzed during the current study is available in the Stanford Network Analysis Project repository, <http://snap.stanford.edu/data/soc-sign-epinions.html>.

Code availability

Source code to simulate opinion formation on synthetic relation networks is available at <https://github.com/8medom/OpinionFormation>.

Received: 5 June 2020; Accepted: 10 March 2021;

Published online: 14 April 2021

References

1. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **124**, 048301 (2020).

2. Del Vicario, M. et al. The spreading of misinformation online. *Proc. Natl Acad. Sci. USA* **113**, 554–559 (2016).
3. Scheufele, D. A. & Krause, N. M. Science audiences, misinformation, and fake news. *Proc. Natl Acad. Sci. USA* **116**, 7662–7669 (2019).
4. Shao, C. et al. The spread of low-credibility content by social bots. *Nat. Commun.* **9**, 1–9 (2018).
5. Stella, M., Ferrara, E. & De Domenico, M. Bots increase exposure to negative and inflammatory content in online social systems. *Proc. Natl Acad. Sci. USA* **115**, 12435–12440 (2018).
6. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
7. Johnson, N. F. et al. The online competition between pro-and anti-vaccination views. *Nature* **582**, 230–233 (2020).
8. Yang, K.-C., Torres-Lugo, C. & Menczer, F. Prevalence of low-credibility information on Twitter during the COVID-19 outbreak. In *Proceedings of the 14th International AAAI Conference on Web and Social Media* <https://doi.org/10.36190/2020.16> (2020) http://workshop-proceedings.icwsm.org/abstract?id=2020_16.
9. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009).
10. Sirbu, A., Loreto, V., Servedio, V. D. & Tria, F. In *Participatory Sensing, Opinions and Collective Awareness* (eds, Loreto, V., Haklay, M., Hotho, A., Servedio, V.D.P., Stumme, G., Theunis, J. & Tria, F.), 363–401 (Springer, 2017) <https://www.springer.com/gp/book/9783319256566>.
11. Schweitzer, F. Sociophysics. *Phys. Today* **71**, 40–46 (2018).
12. Schelling, T. C. Dynamic models of segregation. *J. Math. Sociol.* **1**, 143–186 (1971).
13. DeGroot, M. H. Reaching a consensus. *J. Am. Stat. Assoc.* **69**, 118–121 (1974).
14. Friedkin, N. E. & Johnsen, E. C. Social influence and opinions. *J. Math. Sociol.* **15**, 193–206 (1990).
15. Axelrod, R. The dissemination of culture: a model with local convergence and global polarization. *J. Conflict Resolut.* **41**, 203–226 (1997).
16. Friedkin, N. E. & Bullo, F. How truth wins in opinion dynamics along issue sequences. *Proc. Natl Acad. Sci. USA* **114**, 11380–11385 (2017).
17. Acemoglu, D. & Ozdaglar, A. Opinion dynamics and learning in social networks. *Dyn. Games Appl.* **1**, 3–49 (2011).
18. Easley, D. & Kleinberg, J. *Networks, Crowds, and Markets* (Cambridge University Press, 2010).
19. Newman, M. *Networks* (Oxford University Press, 2018).
20. Lobato, E., Mendoza, J., Sims, V. & Chin, M. Examining the relationship between conspiracy theories, paranormal beliefs, and pseudoscience acceptance among a university population. *Appl. Cogn. Psychol.* **28**, 617–625 (2014).
21. Heider, F. Attitudes and cognitive organization. *J. Psychol.* **21**, 107–112 (1946).
22. Cartwright, D. & Harary, F. Structural balance: a generalization of Heider's theory. *Psychol. Rev.* **63**, 277 (1956).
23. Wasserman, S. & Faust, K. *Social Network Analysis: Methods and Applications*, Vol. 8 (Cambridge University Press, 1994).
24. Moore, M. Structural balance and international relations. *Eur. J. Soc. Psychol.* **9**, 323–326 (1979).
25. Crescenzi, M. J. Reputation and interstate conflict. *Am. J. Political Sci.* **51**, 382–396 (2007).
26. Facchetti, G., Iacono, G. & Altafini, C. Computing global structural balance in large-scale signed social networks. *Proc. Natl Acad. Sci. USA* **108**, 20953–20958 (2011).
27. Zheng, X., Zeng, D. & Wang, F.-Y. Social balance in signed networks. *Inf. Syst. Front.* **17**, 1077–1095 (2015).
28. Lerner, J. & Lomi, A. The free encyclopedia that anyone can dispute: an analysis of the micro-structural dynamics of positive and negative relations in the production of contentious wikipedia articles. *Soc. Netw.* **60**, 11–25 (2020).
29. Adamic, L. A. & Glance, N. The political blogosphere and the 2004 US election: divided they blog. In *Proc. 3rd International Workshop on Link Discovery*, 36–43 (ACM, 2005) <https://dl.acm.org/doi/10.1145/1134271.1134277>.
30. Marvel, S. A., Kleinberg, J., Kleinberg, R. D. & Strogatz, S. H. Continuous-time model of structural balance. *Proc. Natl Acad. Sci. USA* **108**, 1771–1776 (2011).
31. Weigmann, K. The genesis of a conspiracy theory: Why do people believe in scientific conspiracy theories and how do they spread? *EMBO reports* **19**, e45935 (2018).
32. Tversky, A. & Kahneman, D. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* **90**, 293 (1983).
33. Kahneman, D. *Thinking, Fast and Slow* (Macmillan, 2011).
34. Pennycook, G. & Rand, D. G. Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50 (2019).
35. Van de Rijt, A. The micro-macro link for the theory of structural balance. *J. Math. Sociol.* **35**, 94–113 (2011).
36. Harary, F. On the measurement of structural balance. *Behav. Sci.* **4**, 316–323 (1959).
37. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
38. Lorenz, J. Continuous opinion dynamics under bounded confidence: a survey. *Int. J. Mod. Phys. C* **18**, 1819–1838 (2007).
39. Urena, R., Kou, G., Dong, Y., Chiclana, F. & Herrera-Viedma, E. A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. *Inf. Sci.* **478**, 461–475 (2019).
40. Kunegis, J., Lommatzsch, A. & Bauckhage, C. The Slashdot zoo: mining a social network with negative edges. In *Proc. 18th International Conference on World Wide Web*, 741–750 (ACM, 2009) <https://dl.acm.org/doi/proceedings/10.1145/1526709>.
41. Leskovec, J., Huttenlocher, D. & Kleinberg, J. Signed networks in social media. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 1361–1370 (ACM, 2010).
42. Battiston, F., Cairoli, A., Nicosia, V., Baule, A. & Latora, V. Interplay between consensus and coherence in a model of interacting opinions. *Physica D* **323**, 12–19 (2016).
43. Voeten, E., Strezhnev, A. & Bailey, M. United Nations General Assembly Voting Data. <https://doi.org/10.7910/DVN/LEJUQZ> (2020).
44. Bailey, M. A., Strezhnev, A. & Voeten, E. Estimating dynamic state preferences from United Nations voting data. *J. Conflict Resolut.* **61**, 430–456 (2017).
45. Leskovec, J. & Krevl, A. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data> (2014).

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Nos. 11622538, 61673150, 11850410444). M.S.M. acknowledges financial support from the URPP Social Networks at the University of Zurich, the Swiss National Science Foundation (Grant No. 200021-182659), and the UESTC professor research start-up (Grant No. ZYGX2018KYQD215). L.L. acknowledges the Science Strength Promotion Programme of UESTC (Grant No. Y030190261010020).

Author contributions

M.M. conceived the original idea, performed the simulations, and derived the analytical results. M.M., M.S.M., and L.L. interpreted the results and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-021-00579-3>.

Correspondence and requests for materials should be addressed to M.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021